

# MIXING AND TRIPLE RECURRENCE IN PROBABILITY GROUPS

ANUSH TSERUNYAN

**ABSTRACT.** We consider a class of groups equipped with an invariant probability measure (call them probability groups), which includes all compact groups and is closed under taking ultraproducts with the induced Loeb measure. This note develops the basics of the theory of measure-preserving actions of these groups on probability spaces, culminating in a triple recurrence result for mixing probability groups, which generalizes a recent theorem of Bergelson and Tao [BT13] proved for ultra quasirandom groups, nevertheless having a considerably shorter proof<sup>1</sup>. Moreover, the quantitative version of this proof (modifying only the proof of the van der Corput trick using an argument of Austin) yields a quantitative triple recurrence theorem for probability groups that are mixing up to  $\epsilon$  error, such as quasirandom groups introduced by Gowers in [Gow08]. The bound on the error obtained in the latter result is  $4\sqrt{\epsilon}$ , which gives a slight improvement on the bound in the same result obtained by Austin for quasirandom groups in [Aus13].

## 1. PROBABILITY GROUPS AND THEIR ACTIONS

The following definition is suited to incorporate ultraproducts of finite groups equipped with the Loeb measure.

**Definition 1.** Let  $G$  be a group,  $\mathcal{B}, \mathcal{B}_2$  be  $\sigma$ -algebras on  $G$  and  $G^2$ , respectively, so that  $\mathcal{B}_2 \supseteq \mathcal{B} \times \mathcal{B}$ , and  $\mu, \mu_2$  are probability measures on  $\mathcal{B}, \mathcal{B}_2$ , respectively, so that  $\mu_2 \upharpoonright_{\mathcal{B} \times \mathcal{B}} = \mu \times \mu$ . We call the tuple  $(G, \mathcal{B}, \mu, \mathcal{B}_2, \mu_2)$  a **probability group** if

- (i) the multiplication operation  $\cdot : (G^2, \mathcal{B}_2) \rightarrow (G, \mathcal{B})$  is measurable;
- (ii) the inverse operation  $()^{-1} : (G, \mathcal{B}) \rightarrow (G, \mathcal{B})$  is measurable;
- (iii)  $\mu$  is invariant with respect to the left and right translation actions, as well as the inverse operation;
- (iv) for any  $\mathcal{B}_2$ -measurable function  $f : G^2 \rightarrow \mathbb{R}$ , Fubini's theorem holds, i.e.
  - (a) for every  $g \in G$ , the functions  $f_g := f(g, \cdot)$  and  $f^g = f(\cdot, g)$  are  $\mathcal{B}$ -measurable;
  - (b) the functions  $g \mapsto \int_G f(g, h) d\mu(h)$  and  $g \mapsto \int_G f(h, g) d\mu(h)$  are  $\mathcal{B}$ -measurable;
  - (c)  $\int_G \int_G f(g, h) d\mu(g) d\mu(h) = \int_{G^2} f(g, h) d\mu_2(g, h) = \int_G \int_G f(g, h) d\mu(h) d\mu(g)$ .

The author was surprised to also find this definition in [Wei65] as it doesn't seem like Weil applies it to ultraproducts, which is where having a stronger  $\sigma$ -algebra on the product is needed; so it is possible that this paper is the first concrete application of the full power of the above definition. We refer to [Wei65] for a systematic study of probability groups in general (although Weil doesn't give them a name).

**Remark 2.** Clause (iv)(a) of the above definition implies that for every  $g \in G$ , the function  $f_g : (G, \mathcal{B}) \rightarrow (G^2, \mathcal{B}_2)$  defined by  $f_g(h) = (g, h)$  is measurable: indeed, for any  $A \in \mathcal{B}_2$ , (iv)(a) applied to the characteristic function  $\chi_A : (G^2, \mathcal{B}_2) \rightarrow \mathbb{R}$  gives that for every  $g \in G$ , the fiber  $A_g = f_g^{-1}(A)$  is in  $\mathcal{B}$ . Similarly, the function  $f^g : (G, \mathcal{B}) \rightarrow (G^2, \mathcal{B}_2)$  defined by  $f^g(h) = (g, h)$  is also measurable. This implies that for each  $g \in G$ , the functions of left and right multiplications by  $g$  are measurable.

<sup>1</sup>Similar results were obtained independently by Tim Austin in [Aus13].

Below we omit writing the  $\sigma$ -algebras  $\mathcal{B}, \mathcal{B}_2$  and the measure  $\mu_2$  if they are understood or irrelevant.

### Examples.

- (3.A) Finite groups with normalized counting measure are probability groups.
- (3.B) More generally, any compact Hausdorff group with its normalized Haar measure is a probability group (in this case  $\mathcal{B}_2 = \mathcal{B}^2$ ).
- (3.C) Ultraproduct of compact Hausdorff groups is a probability group with respect to the induced Loeb  $\sigma$ -algebras and Loeb measures. We refer to Proposition 4 for a more precise statement, as well as to [BT13] and [CKTD12] for nice expositions of the ultraproduct and Loeb measure constructions.
- (3.D) Ultraproduct of amenable groups is a probability group with respect to the induced Loeb  $\sigma$ -algebras and Loeb measures; more precisely, if each  $G_n$  is an amenable group equipped with a *finitely additive* invariant probability measure  $\mu_n$ , then the Loeb measure  $\mu$  of the ultraproduct  $(G, \mu)$  of the sequence  $(G_n, \mu_n)_{n \in \mathbb{N}}$  is actually *countably additive*, making  $(G, \mu)$  a probability group.

In general, we have:

**Proposition 4.** *Ultraproduct of probability groups together with the induced Loeb measure is a probability group.*

*Proof.* Let  $\alpha$  be an ultrafilter on  $\mathbb{N}$  and  $(G_n, \mathcal{B}^{(n)}, \mu^{(n)}, \mathcal{B}_2^{(n)}, \mu_2^{(n)})_{n \in \mathbb{N}}$  be a sequence of probability groups. Let  $G = \prod_{n \rightarrow \alpha} G_n$  and let  $\mathcal{B}, \mathcal{B}_2$  be the  $\sigma$ -algebras generated by  $\prod_{n \rightarrow \alpha} \mathcal{B}^{(n)}$  and  $\prod_{n \rightarrow \alpha} \mathcal{B}_2^{(n)}$  on  $G$  and  $G^2$ , respectively. Lastly, let  $\mu, \mu_2$  be the Loeb measures on  $G$  and  $G^2$  induced by  $\prod_{n \rightarrow \alpha} \mu^{(n)}$  and  $\prod_{n \rightarrow \alpha} \mu_2^{(n)}$ . Now it is not hard to check that  $(G, \mathcal{B}, \mu, \mathcal{B}_2, \mu_2)$  is a probability group and we refer to Theorem 19 of [BT13] for verification of Fubini's theorem (although in [BT13]  $G_n$  are assumed to be finite, the argument works equally well for general probability groups).  $\square$

**Remark 5.** When taking an ultraproduct  $(G, \mathcal{B}) = \prod_{n \rightarrow \alpha} (G_n, \mathcal{B}^{(n)})$ , even when  $G_n$  are finite and  $\mathcal{B}^{(n)} = \text{Pow}(G_n)$ , the group multiplication on  $G$  may not be measurable with respect to  $\mathcal{B} \times \mathcal{B}$ . This is why it is necessary to have a stronger  $\sigma$ -algebra  $\mathcal{B}_2$  on  $G \times G$ , namely the one generated by  $\prod_{n \rightarrow \alpha} \mathcal{B}^{(n)} \times \mathcal{B}^{(n)}$ .

We now define a natural class of actions for probability groups.

**Definition 6.** *Let  $(G, \mathcal{B}, \mu, \mathcal{B}_2, \mu_2)$  be a probability group,  $(X, \mathcal{C}, \nu)$  a probability space, and let  $a : G \times X \rightarrow X$  be an action of  $G$  on  $X$ , i.e.  $a(g, x) = g \cdot_a x$ . We would call this action **measure-preserving** if there is a  $\sigma$ -algebra  $\mathcal{C}_2 \supseteq \mathcal{B} \times \mathcal{C}$  on  $G \times X$  such that*

- (i) *the action is measurable as a function  $a : (G \times X, \mathcal{C}_2) \rightarrow (X, \mathcal{C})$ .*
- (ii) *the action preserves the measure  $\nu$ , i.e.  $\nu(g^{-1} \cdot_a A) = \nu(A)$  for all  $g \in G$  and  $A \in \mathcal{C}$ ;*
- (iii) *for any  $\mathcal{C}_2$ -measurable function  $f : G \times X \rightarrow \mathbb{R}$ , Fubini's theorem holds (as in the definition of a probability group).*

**Remark 7.** Just like in the definition of a probability group, here also clause (iii) implies that for every  $g \in G$  and  $x \in X$ , the functions  $f_g : (X, \mathcal{C}) \rightarrow (G \times X, \mathcal{C}_2)$  defined by  $f_g(x') = (g, x')$

and  $f^x : (G, \mathcal{B}) \rightarrow (G \times X, \mathcal{C}_2)$  defined by  $f^x(g') = (g')$  are measurable. In particular, the action of every fixed  $g \in G$  on  $X$  is measurable.

**Examples 8.** For a probability group  $G$ , the left and right translation actions  $g \cdot_l x \mapsto gx$  and  $g \cdot_r x \mapsto xg^{-1}$ , as well as the conjugation action  $g \cdot_c x \mapsto gxg^{-1}$  of  $G$  on itself, are measure-preserving (with  $\mathcal{C}_2 = \mathcal{B}_2$ ).

**Unitary representations 9.** A measure-preserving action  $a : G \curvearrowright (X, \nu)$  as above induces an action  $G \curvearrowright L^2(X, \nu)$ , still denoted by  $\cdot_a$  and defined by

$$(g \cdot_a f)(x) = f(g^{-1} \cdot_a x).$$

In fact this action is unitary and is known as the Koopman representation. Let  $\text{Inv}_a(X, \nu) \subseteq L^2(X, \nu)$  denote the subspace of functions  $f$  invariant under this action (i.e.  $g \cdot_a f = f$  for all  $g \in G$ ). Finally, let  $P_a : L^2(X, \nu) \rightarrow \text{Inv}_a(X, \nu)$  be the orthogonal projection onto  $\text{Inv}_a(X, \nu)$ . Below we use  $\langle \cdot, \cdot \rangle_X$  to denote the inner product in  $L^2(X, \nu)$ . All  $L^2$ -spaces and Hilbert spaces in general are assumed to be complex.

If  $G$  is a probability group and the action  $a : G \curvearrowright G$  is either the left or right translation, then for  $f \in L^2(G)$ ,  $P_a(f)$  is just the mean of  $f$  because these actions are transitive, so the only invariant functions are constants. In general, here is how to explicitly compute  $P_a$  for arbitrary measure-preserving actions.

**Proposition 10** (Mean ergodic theorem for probability groups). *Let  $(G, \mu)$  be a probability group,  $(X, \nu)$  a probability space and let  $a : G \curvearrowright X$  be a measure-preserving action. Then for all  $f \in L^2(X, \nu)$ ,*

$$P_a(f)(x) = \int_G (g \cdot_a f)(x) d\mu(g).$$

*In particular, if the action is ergodic (i.e. any measurable invariant subset of  $X$  is either  $\nu$ -null or  $\nu$ -conull), then for  $\nu$ -a.e.  $x \in X$ ,*

$$\int_G (g \cdot_a f)(x) d\mu(g) = \int_X f(y) d\nu(y).$$

*Proof.* Fix  $\phi \in \text{Inv}_a(X, \nu)$ ; we need to show that  $f - \int_G (g \cdot_a f)(x) d\mu(g)$  and  $\phi$  are orthogonal, for which it is enough to show that  $\langle f, \phi \rangle_X = \langle \int_G (g \cdot_a f)(x) d\mu(g), \phi \rangle_X$ . Compute:

$$\begin{aligned} \langle \int_G (g \cdot_a f)(x) d\mu(g), \phi \rangle_X &= \int_X \int_G (g \cdot_a f)(x) \phi(x) d\mu(g) d\nu(x) \\ [\text{Fubini}] &= \int_G \int_X (g \cdot_a f)(x) \phi(x) d\nu(x) d\mu(g) \\ [\text{replace variable } x \text{ with } g^{-1} \cdot_a x] &= \int_G \int_X f(x) (g^{-1} \cdot_a \phi)(x) d\nu(x) d\mu(g) \\ [\text{invariance of } \phi] &= \int_G \int_X f(x) \phi(x) d\nu(x) d\mu(g) \\ &= \int_G \langle f, \phi \rangle_X d\mu(g) = \langle f, \phi \rangle_X. \end{aligned}$$

Furthermore, if the action is ergodic, then the only functions in  $\text{Inv}_a(X, \nu)$  are constants, so  $P_a(f) = \int_X f(x) d\nu(x)$ .  $\square$

## 2. MIXING FOR PROBABILITY GROUPS

For a measure  $\mu$ , we write  $\forall^\mu$  to mean “for  $\mu$ -a.e.”.

**Definition 11.** Let  $a : G \curvearrowright X$  be a measure-preserving action of a probability group  $(G, \mu)$  on a probability space  $(X, \nu)$ . Call this action **mixing along  $\mu$**  (or just **mixing**) if for any  $f_1, f_2 \in L^2(X, \nu)$ ,

$$(\forall^\mu g \in G) \langle f_1, g \cdot_a f_2 \rangle_X = \langle P_a(f_1), P_a(f_2) \rangle_X.$$

One could also give an abstract definition of *mixing along a filter*  $\mathcal{F} \subseteq \text{Pow}(G)$  for any group  $G$  as follows: for any  $f_1, f_2 \in L^2(X, \nu)$ ,

$$\lim_{g \rightarrow \mathcal{F}} \langle f_1, g \cdot_a f_2 \rangle_X = \langle P_a(f_1), P_a(f_2) \rangle_X.$$

For ergodic actions, this generalizes the usual notions of mixing such as

- *weak mixing* for amenable  $G$  and filter  $\mathcal{F}$  of density-one sets;
- *mild mixing* for arbitrary discrete  $G$  and filter  $\mathcal{F} = \text{IP}^*$ ;
- *strong mixing* for arbitrary discrete  $G$  and the Fréchet filter  $\mathcal{F}$ .

In our case, due to the countable additivity of  $\mu$ , the definition of  $\mu$ -mixing is equivalent to mixing along the filter of  $\mu$ -conull sets.

**Remark.** A similar definition of mixing along a filter for ergodic actions was considered by Tucker-Drob in Chapter 7 of [TD13].

**Example 12** (Ultra quasirandom groups). In [BT13], the authors consider finite groups that are approximately mixing (i.e. mixing with a small error); more precisely, they consider so-called *D-quasirandom groups*, introduced by Gowers in [Gow08], which are finite (or more generally compact Hausdorff) groups that do not admit any nontrivial unitary representations of dimension  $< D$ . It is then shown that the right translation action of these groups on themselves is mixing with an error  $D^{-1/2}$ , with respect to the normalized Haar measure (see Proposition 3 in [BT13] or the last section of the current paper). Therefore, taking an appropriate ultraproduct washes the error away, yielding a probability group whose right translation action on itself is genuinely mixing. More precisely, Bergelson-Tao define *ultra quasirandom groups* as an ultraproduct of a sequence  $(G_n, \mu_n)_n$  of finite groups, where  $\mu_n$  is the normalized counting measure, each  $G_n$  is  $D_n$ -quasirandom and  $D_n \rightarrow \infty$ . This is a probability group with respect to the induced Loeb measure, and, by Lemma 33 in [BT13], its right translation action on itself is mixing.

We are finally ready to give the main definition, which at a glance may seem hard to check and unlikely to occur, but Proposition 14 below will settle the matter.

**Definition 13.** We call a probability group **mixing** if all of its measure-preserving actions on probability spaces are mixing.

**Proposition 14.** A probability group  $(G, \mu)$  is mixing if and only if its right translation action on itself is mixing.

*Proof.* We show the nontrivial direction: suppose the right translation action  $r : G \curvearrowright G$  is mixing and consider a measure-preserving action  $a : G \curvearrowright X$  on a probability space  $(X, \nu)$ .

The idea is to switch from averaging over the action  $a : G \curvearrowright X$  to averaging over the right translation action  $r : G \curvearrowright G$ ; this is done using Fubini’s theorem and the following trivial identity (associativity of the action): for  $g, h \in G$  and  $x \in X$ ,

$$g \cdot_a (h \cdot_a x) = (gh) \cdot_a x = (h \cdot_r g) \cdot_a x.$$

Turning to the actual proof, for a function  $f : X \rightarrow \mathbb{R}$  and  $x \in X$ , let  $f^{(x)} : G \rightarrow \mathbb{R}$  be defined by  $f^{(x)}(g) = g \cdot_a f(x)$ . Now fix  $f_1, f_2 \in L^2(X, \nu)$ , and for  $g \in G$ , compute:

$$\begin{aligned}
\langle f_1, g \cdot_a f_2 \rangle_X &= \int_X f_1(x)(g \cdot_a f_2(x))d\nu(x) \\
[\text{dummy integration over } G] &= \int_G \int_X f_1(x)(g \cdot_a f_2(x))d\nu(x)d\mu(h) \\
[\text{replace variable } x \text{ with } h^{-1} \cdot_a x] &= \int_G \int_X (h \cdot_a f_1(x))(h \cdot_a g \cdot_a f_2(x))d\nu(x)d\mu(h) \\
[\text{Fubini}] &= \int_X \int_G f_1^{(x)}(h)f_2^{(x)}(hg)d\mu(h)d\nu(x). \\
&= \int_X \langle f_1^{(x)}, g \cdot_r f_2^{(x)} \rangle_G d\nu(x).
\end{aligned}$$

Because the right translation action is mixing and ergodic, we have

$$(\forall x \in X)(\forall^\mu g \in G) \langle f_1^{(x)}, g \cdot_r f_2^{(x)} \rangle_G = \left( \int_G f_1^{(x)} \right) \left( \int_G f_2^{(x)} \right).$$

Thus, Fubini's theorem gives

$$(\forall^\mu g \in G)(\forall^\nu x \in X) \langle f_1^{(x)}, g \cdot_r f_2^{(x)} \rangle_G = \left( \int_G f_1^{(x)} \right) \left( \int_G f_2^{(x)} \right).$$

Noting that by the mean ergodic theorem (Proposition 10),  $\int_G f^{(x)} = P_a(f)(x)$  for  $f \in L^2(X, \nu)$ , we get

$$(\forall^\mu g \in G) \langle f_1, g \cdot_a f_2 \rangle_X = \int_X P_a(f_1)(x)P_a(f_2)(x)d\mu(x) = \langle P_a(f_1), P_a(f_2) \rangle_X.$$

□

**Example 15.** As mentioned in Example 12, the right translation action of an ultra quasirandom group on itself is mixing. Thus, ultra quasirandom groups are mixing probability groups. This, in particular, implies Lemma 34 of [BT13].

### 3. TRIPLE RECURRENCE FOR MIXING PROBABILITY GROUPS

We now state the main result of the paper, namely, the fact that in probability groups, mixing (i.e. double recurrence) can be amplified to a triple recurrence. This generalizes Theorem 41 in [BT13] proven for ultra quasirandom groups.

**Theorem 16.** *Let  $(G, \mu)$  be a mixing probability group. Then for any  $f_1, f_2, f_3 \in L^\infty(G, \mu)$ ,*

$$(\forall^\mu g \in G) \int_G f_1(x)(g \cdot_l f_2)(x)(g \cdot_c f_3)(x)dx = \int_G f_1(x)P_l(f_2)(x)P_c(f_3)(x)dx,$$

where  $\cdot_l$  and  $\cdot_c$  are, respectively, the left translation and the conjugation actions of  $G$  on itself.

Using transfer principle (or equivalently, considering an ultraproduct of counter-example quasirandom groups with  $D \rightarrow \infty$ ), Bergelson and Tao show in [BT13, Theorem 5] that this theorem for ultra quasirandom groups implies an analogous quantitative triple recurrence result for finite quasirandom groups with an implicit bound on error. See [BT13, Corollary 7] for the interpretation of this result in terms of the distribution of the quadruples  $(g, x, gx, xg)$  with  $x, g$  drawn uniformly and independently at random. See also [BT13, Corollary 8] for a density noncommutative Schur theorem for quasirandom groups.

**The idea of the proof.** If we remove one of the factors  $f_1$ ,  $g \cdot_l f_2$  or  $g \cdot_c f_3$  from the desired equality, i.e. “drop the degree” of the product, then the equality would easily follow from double recurrence, i.e. the fact that  $G$  is mixing. So we get rid of the factor  $f_1$  and here is how. Linearity reduces to the orthogonal cases  $P_c(f_3) = f_3$  and  $P_c(f_3) = 0$ , and the proof for the former case falls out of left translation action being mixing, so we are left with the case  $P_c(f_3) = 0$ . Assuming this, what we need to show is

$$\forall^\mu g \langle f_1, e_g \rangle_G = 0,$$

where  $e_g = (g \cdot_l f_2)(g \cdot_c f_3)$ . But the latter would follow basically from Bessel’s inequality if we could show that  $\{e_g\}_{g \in G}$  is an a.e.-orthogonal family in  $L^2(G, \mu)$ , i.e.

$$\forall^{\mu^2} (g, h) \langle e_g, e_h \rangle_G = 0.$$

By Fubini’s theorem and a change of variable, this is equivalent to

$$\forall^\mu h \forall^\mu g \langle e_g, e_{gh} \rangle_G = 0,$$

which, due to some regrouping and cancellation, easily follows from the right translation and the conjugation actions being mixing. This latter trick of replacing pairs  $(g, h)$  by  $(g, gh)$  is known as the *van der Corput difference trick*, which can be thought of as an analog of differentiation in this context because an application of this trick “drops the degree”.

**Remark.** In the proof of this theorem for an ultra quasirandom group given in [BT13], the authors restrict to a countable subgroup  $\Gamma$  of  $G$  and use an idempotent ultrafilter on  $\Gamma$  as their notion of largeness that is almost invariant under the translation action of  $\Gamma$  on itself. We instead use the measure  $\mu$  on  $G$ , or equivalently, the filter of  $\mu$ -conull sets, which is genuinely invariant and also has the advantage of being countably additive; the latter enables cleaner pigeon-hole arguments and replaces various limits with a.e. statements. The only price we pay is that our filter of  $\mu$ -conull sets is not “ultra”, but this is not an issue as we can be careful enough to stay in the  $\sigma$ -algebra of measurable sets when needed.

#### 4. PROOF OF TRIPLE RECURRENCE

We start by recording a (cheap) Ramsey theorem for filters. For a filter  $\mathcal{F}$  on a set  $X$ , we write  $\forall^\mathcal{F}$  below to mean “for an  $\mathcal{F}$ -large set of points in  $X$ ”.

**Lemma 17** (Ramsey for filters). *Let  $X$  be a set and  $\mathcal{F}$  a nonprincipal filter on it. Let  $R \subseteq X^2$  be such that*

$$(\forall^\mathcal{F} x \in X) (\forall^\mathcal{F} y \in X) x R y.$$

*Then there is an infinite set  $\{x_n\}_{n \in \mathbb{N}} \subseteq X$  such that  $x_n R x_m$  for all  $n < m$ .*

*Proof.* By the hypothesis,  $A = \{x \in X : R_x \text{ is } \mathcal{F}\text{-large}\}$  is  $\mathcal{F}$ -large. Put  $A_0 = A$  and take  $x_0 \in A_0$ . Put  $A_1 = R_{x_0} \cap A_0$  and note that  $A_1$  is still  $\mathcal{F}$ -large. Take  $x_1 \in A_1$  distinct from  $x_0$  (can do this because  $\mathcal{F}$  is nonprincipal). Repeat: put  $A_2 = R_{x_1} \cap A_1$  and note that  $A_2$  is still  $\mathcal{F}$ -large. Take  $x_2 \in A_2$  distinct from  $x_0, x_1$ ; etc.  $\square$

Now recall the following basic Hilbert space fact, whose proof is immediate from Bessel’s inequality:

**Lemma 18** (Bessel). *Let  $(e_n)_{n \in \mathbb{N}}$  be a bounded sequence of vectors in a Hilbert space  $\mathcal{H}$ . If the vectors in  $(e_n)_{n \in \mathbb{N}}$  are pairwise orthogonal, then  $\lim_{n \rightarrow \infty} e_n = 0$  in the weak topology of  $\mathcal{H}$ , i.e. for every  $f \in \mathcal{H}$ ,  $\lim_{n \rightarrow \infty} \langle f, e_n \rangle = 0$ .*



Putting this together with the Ramsey lemma for the filter of conull sets, we get a natural analog of Bessel's lemma for measure:

**Lemma 19** (Random Bessel). *Let  $(X, \mu)$  be a measure space with nonatomic  $\mu \neq 0$  and let  $(e_x)_{x \in X}$  be a bounded sequence in a Hilbert space  $\mathcal{H}$ . If*

$$(\forall^\mu x \in X) (\forall^\mu y \in X) \langle e_x, e_y \rangle = 0,$$

*then for every  $f \in \mathcal{H}$ ,  $(\forall^\mu x \in X) \langle f, e_x \rangle = 0$ .*

*Proof.* Fix  $f \in \mathcal{H}$  and suppose that the conclusion fails for this  $f$ . Then, there is  $\epsilon > 0$  such that the set  $Y = \{x \in X : |\langle f, e_x \rangle| \geq \epsilon\}$  is not  $\mu$ -null (caution:  $Y$  may not be measurable). Thus, the restriction of the filter of  $\mu$ -conull sets to  $Y$  gives a nonprincipal filter  $\mathcal{F}$  on  $Y$ . Applying the Ramsey Lemma 17 to  $Y$  with filter  $\mathcal{F}$  and  $R = \{(x, y) \in Y^2 : \langle e_x, e_y \rangle = 0\}$ , we get an infinite bounded sequence  $(e_{x_n})_{n \in \mathbb{N}}$  of pairwise orthogonal vectors such that for every  $n \in \mathbb{N}$ ,  $|\langle f, e_{x_n} \rangle| \geq \epsilon$ , contradicting Lemma 18.  $\square$

Inviting group structure and Fubini to this party of Ramsey and Bessel, we get:

**Lemma 20** (Random van der Corput). *Let  $(G, \mathcal{B}, \mu, \mathcal{B}_2, \mu_2)$  be an infinite probability group and let  $(e_g)_{g \in G}$  be a bounded sequence in a Hilbert space  $\mathcal{H}$  such that the function  $(g, h) \mapsto \langle e_g, e_h \rangle$  is measurable as a function  $(G^2, \mathcal{B}_2) \rightarrow \mathbb{R}$ . If*

$$(\forall^\mu h \in G) (\forall^\mu g \in G) \langle e_g, e_{gh} \rangle = 0,$$

*then for all  $f \in \mathcal{H}$ ,  $(\forall^\mu g \in G) \langle f, e_g \rangle = 0$ .*

*Proof.* Applying Fubini's theorem, we get  $\forall^\mu g \forall^\mu h \langle e_g, e_{gh} \rangle = 0$ . Changing the variable  $h \mapsto gh$ , we fulfill the hypothesis of Lemma 19 and the conclusion follows.  $\square$

**Remark.** This lemma has several cousins in the countable setting; e.g. for the filter on  $\mathbb{N}$  of sets of density 1 (Lemma 4.9 in [Fur81]), for the filter on  $\mathbb{N}$  of sets that meet every IP-set (Lemma 9.24 in [Fur81]) and for idempotent ultrafilters on countable groups (Theorem 2.3 in [BM07]). See also Lemma 23 for a quantitative version.

We are now ready to prove the triple recurrence theorem.

*Proof of Theorem 16.* As we solely work in  $G$ , we write  $\langle \cdot, \cdot \rangle$  instead of  $\langle \cdot, \cdot \rangle_G$ .

Since  $g \cdot_c P_c(f_3) = P_c(f_3)$  and hence

$$\langle f_1(g \cdot_l f_2), g \cdot_c f_3 \rangle = \langle f_1(g \cdot_l f_2), g \cdot_c (f_3 - P_c(f_3)) \rangle + \langle f_1(g \cdot_l f_2), P_c(f_3) \rangle,$$

it is enough to prove the theorem in the following two orthogonal cases:

*Case  $P_c(f_3) = f_3$ :* In this case, what we need to show is

$$(\forall^\mu g \in G) \langle f_1 f_3, g \cdot_l f_2 \rangle = \langle f_1 f_3, P_l(f_2) \rangle,$$

which immediately follows from the fact that the left translation action is mixing.

*Case  $P_c(f_3) = 0$ :* Now what we need to show is

$$(\forall^\mu g \in G) \langle f_1, (g \cdot_l f_2)(g \cdot_c f_3) \rangle = 0,$$

which will follow from the random van der Corput lemma for  $e_g = (g \cdot_l f_2)(g \cdot_c f_3)$  once we verify hypothesis. It easily follows from Fubini's theorem and the definitions that the function  $(G^2, \mathcal{B}_2) \rightarrow \mathbb{R}$  defined by

$$(g, h) \mapsto \langle e_g, e_h \rangle = \int_G f_2(g^{-1} \cdot_l x) f_3(g^{-1} \cdot_c x) f_2(h^{-1} \cdot_l x) f_3(h^{-1} \cdot_c x) dx$$

is measurable. Furthermore, the family  $\{e_g\}_{g \in G}$  in  $L^2(X, \mu)$  is bounded because  $f_2, f_3 \in L^\infty(G, \mu)$  and  $\mu$  is finite. It remains to verify that  $\forall^\mu h \forall^\mu g \langle e_g, e_{gh} \rangle = 0$ . To this end, fix  $h, g \in G$  and compute:

$$\begin{aligned}
\langle e_g, e_{gh} \rangle &= \int_G (g \cdot_l f_2)(g \cdot_c f_3)((gh) \cdot_l f_2)((gh) \cdot_c f_3) dx \\
[\text{associativity of actions and regrouping}] &= \int_G [(g \cdot_l f_2)(g \cdot_l h \cdot_l f_2)][(g \cdot_c f_3)(g \cdot_c h \cdot_c f_3)] dx \\
[\text{distributivity of actions over product}] &= \int_G [g \cdot_l (f_2(h \cdot_l f_2))][g \cdot_c (f_3(h \cdot_c f_3))] dx \\
\left[ \begin{array}{l} F_2^{(h)} := f_2(h \cdot_l f_2) \\ F_3^{(h)} := f_3(h \cdot_c f_3) \end{array} \right] &= \int_G [g \cdot_l F_2^{(h)}][g \cdot_c F_3^{(h)}] dx \\
[g \cdot_c f = g \cdot_l g \cdot_r f] &= \int_G [g \cdot_l F_2^{(h)}][g \cdot_l g \cdot_r F_3^{(h)}] dx \\
[\text{distributivity of actions over product}] &= \int_G g \cdot_l [F_2^{(h)}(g \cdot_r F_3^{(h)})] dx \\
[\text{replace variable } x \text{ with } g \cdot_l x] &= \int_G F_2^{(h)}(g \cdot_r F_3^{(h)}) dx = \langle F_2^{(h)}, g \cdot_r F_3^{(h)} \rangle.
\end{aligned}$$

Because the right translation action is mixing, we have that for every  $h \in G$ :

$$(\forall^\mu g) \langle F_2^{(h)}, g \cdot_r F_3^{(h)} \rangle = \left( \int_G F_2^{(h)} \right) \left( \int_G F_3^{(h)} \right).$$

But the conjugation action is mixing as well, so we get

$$(\forall^\mu h) \int_G F_3^{(h)} = \int_G f_3(h \cdot_c f_3) = \int_G f_3 P_c(f_3) = 0.$$

Thus, we finally have

$$(\forall^\mu h \forall^\mu g) \langle e_g, e_{gh} \rangle = \left( \int_G F_2^{(h)} \right) \left( \int_G F_3^{(h)} \right) = \left( \int_G F_2^{(h)} \right) \cdot 0 = 0.$$

□

## 5. A QUANTITATIVE VERSION

We now work out a quantitative version of the triple recurrence theorem, where we consider probability groups that may not be purely mixing, but are mixing with some error (called  $\epsilon$ -mixing below).

**Credits.** The argument below is the same as above for the infinitary version (replacing the a.e. statements with averages), except for the proof of the quantitative van der Corput lemma (Lemma 23). The proof of the infinitary version of this lemma (Lemma 20) uses a Ramsey-theoretic argument, so the bound obtained from its quantitative version is quite rough and messy to compute. Thus, in the original version of the current paper, quantitative triple recurrence was only mentioned in a remark with its proof omitted because the bound it gave was superseded by [Aus13, Theorem 1], where a nice bound of  $4D^{-1/8}$  was obtained for  $D$ -quasirandom groups. However, after receiving the original version of the current paper (private communication), Tim Austin pointed out an argument replacing the Ramsey-theoretic part of the proof with applications of Fubini's theorem and Cauchy-Schwarz. With Tim's permission, we use this latter argument in the proof of the quantitative van der Corput lemma below and obtain a slightly better bound of  $4D^{-1/4}$  for the triple recurrence theorem.

The exposition below is mainly self-contained and, although written for probability groups, the main application we have in mind is to the following class of groups:



**Definition 21** (Gowers [Gow08]). For  $D \geq 1$ , a compact Hausdorff group  $G$  is called  $D$ -quasirandom if it does not admit any nontrivial unitary representations of dimension  $< D$ . We will always equip  $G$  with its normalized Haar measure, and write  $(G, \mu)$ , turning  $G$  into a probability group.

**Definition 22.** For  $\epsilon > 0$ , call a measure-preserving action  $a : G \curvearrowright X$  of a probability group  $(G, \mu)$  on a probability space  $(X, \nu)$   $\epsilon$ -**mixing** if for any  $f_1, f_2 \in L^2(X, \nu)$ ,

$$\int_G |\langle f_1, g \cdot_a f_2 \rangle - \langle P_a(f_1), P_a(f_2) \rangle| d\mu(g) \leq \epsilon \|f_1\|_{L^2} \|f_2\|_{L^2},$$

where  $\langle \cdot, \cdot \rangle$  denotes the  $L^2$ -inner product. Furthermore, call a probability group  $G$   $\epsilon$ -mixing if so are all of its measure-preserving actions.

[BT13, Proposition 3], as written, states that the right translation action of a  $D$ -quasirandom group on itself is  $D^{-1/2}$ -mixing, but running its proof for any other measure-preserving action yields that  $D$ -quasirandom groups are  $D^{-1/2}$ -mixing.

**Quantitative van der Corput Lemma 23.** Let  $(G, \mathcal{B}, \mu, \mathcal{B}_2, \mu_2)$  be a probability group and let  $(e_g)_{g \in G}$  be a bounded (in the  $L^2$ -norm) sequence in  $L^2(X, \nu)$ , for some probability space  $(X, \nu)$ , such that

- (i) the function  $(g, h) \mapsto \langle e_g, e_h \rangle$  is measurable as a function  $(G^2, \mathcal{B}_2) \rightarrow \mathbb{R}$ ;
- (ii) for every  $f \in L^2(X, \nu)$ , the function  $g \mapsto \langle f, e_g \rangle$  is measurable as a function  $(G, \mathcal{B}) \rightarrow \mathbb{R}$ .

For  $\epsilon \geq 0$ , if

$$\int_G \int_G |\langle e_g, e_{gh} \rangle| d\mu(g) d\mu(h) \leq \epsilon,$$

then for all  $f \in L^2(X, \nu)$ ,

$$\int_G |\langle f, e_g \rangle| d\mu(g) \leq \sqrt{\epsilon} \|f\|_{L^2}.$$

*Proof* (Austin). Let  $\varphi : G \rightarrow \mathbb{C}$  be defined so that  $|\langle f, e_g \rangle| = \varphi(g) \langle f, e_g \rangle$ . Then

$$\begin{aligned} \int_G |\langle f, e_g \rangle| d\mu(g) &= \int_G \int_X \varphi(g) f(x) e_g(x) d\nu(x) d\mu(g) \\ [\text{Fubini}] &= \int_X f(x) \int_G \varphi(g) e_g(x) d\mu(g) d\nu(x) \\ [\text{Cauchy-Schwarz}] &\leq \|f\|_{L^2} \left\| \int_G \varphi(g) e_g(x) d\mu(g) \right\|_{L^2}. \end{aligned}$$

But the second factor in the last term is bounded by  $\sqrt{\epsilon}$  as the following calculation shows:

$$\begin{aligned} \left\| \int_G \varphi(g) e_g(x) d\mu(g) \right\|_{L^2}^2 &= \int_X \int_G \int_G \varphi(g) \overline{\varphi(h)} e_g(x) \overline{e_h(x)} d\mu(h) d\mu(g) d\nu(x) \\ [\text{replace variable } h \text{ with } gh] &= \int_X \int_G \int_G \varphi(g) \overline{\varphi(gh)} e_g(x) \overline{e_{gh}(x)} d\mu(h) d\mu(g) d\nu(x) \\ [\text{Fubini}] &= \int_G \int_G \varphi(g) \overline{\varphi(gh)} \langle e_g, e_{gh} \rangle d\mu(g) d\mu(h) \\ [\text{triangle inequality}] &\leq \int_G \int_G |\langle e_g, e_{gh} \rangle| d\mu(g) d\mu(h) \leq \epsilon. \end{aligned}$$

□

**Theorem 24.** *Let  $0 \leq \epsilon \leq 1$  and let  $(G, \mu)$  be an  $\epsilon$ -mixing probability group. Then for any  $f_1, f_2, f_3 \in L^2(G, \mu)$  with  $\|f_1\|_{L^\infty}, \|f_2\|_{L^\infty}, \|f_3\|_{L^\infty} \leq 1$ ,*

$$\int_G \left| \int_G f_1(x)(g \cdot_l f_2)(x)(g \cdot_c f_3)(x) dx - \int_G f_1(x) P_l(f_2)(x) P_c(f_3)(x) dx \right| dg \leq \epsilon + \sqrt{5\epsilon} \leq 4\sqrt{\epsilon}.$$

*Proof.* Note that  $\|P_c(f_3)\|_{L^\infty} \leq 1$  (by Proposition 10), so  $\|f_3 - P_c(f_3)\|_{L^\infty} \leq 2$ ; however, by Pythagorean theorem, we still have  $\|f_3 - P_c(f_3)\|_{L^2} \leq \|f_3\|_{L^2} \leq \|f_3\|_{L^\infty} \leq 1$ . Thus, the orthogonal decomposition  $f_3 = P_c(f_3) + (f_3 - P_c(f_3))$ , together with triangle inequality, reduces proving the desired inequality to proving the following two inequalities with orthogonal assumptions on  $f_3$ :

(i) assuming  $P_c(f_3) = f_3$  and  $\|f_3\|_{L^\infty} \leq 1$ ,

$$\int_G \left| \int_G f_1(x) f_3(x) (g \cdot_l f_2)(x) dx - \int_G f_1(x) f_3(x) P_l(f_2)(x) dx \right| dg \leq \epsilon;$$

(ii) assuming  $P_c(f_3) = 0$  and  $\|f_3\|_{L^\infty} \leq 2, \|f_3\|_{L^2} \leq 1$ ,

$$\int_G \left| \int_G f_1(x) (g \cdot_l f_2)(x) (g \cdot_c f_3)(x) dx \right| dg \leq \sqrt{5\epsilon}.$$

Case (i) is just the statement of  $\epsilon$ -mixing of the left translation action applied to functions  $f_1 f_3$  and  $f_2$ , so we focus on Case (ii) now. Thus, suppose  $P_c(f_3) = 0$  and put  $e_g = (g \cdot_l f_2)(g \cdot_c f_3)$ . The quantitative van der Corput lemma reduces proving the inequality in Case (ii) to the following:

$$\int_G \int_G |\langle e_g, e_{gh} \rangle| dg dh \leq 5\epsilon,$$

where  $\langle \cdot, \cdot \rangle$  denotes the inner product in  $L^2(G, \mu)$ . The computation done in the proof of Theorem 16 (algebraic manipulations followed by a change of variable) gives:

$$\langle e_g, e_{gh} \rangle = \langle F_2^{(h)}, g \cdot_r F_3^{(h)} \rangle,$$

where  $F_2^{(h)} := f_2(h \cdot_l f_2)$  and  $F_3^{(h)} := f_3(h \cdot_c f_3)$ . So for fixed  $h$ , we get:

$$\begin{aligned} \int_G |\langle e_g, e_{gh} \rangle| dg &= \int_G |\langle F_2^{(h)}, g \cdot_r F_3^{(h)} \rangle| dg \\ [\text{triangle inequality}] &\leq \int_G |\langle F_2^{(h)}, g \cdot_r F_3^{(h)} \rangle - \langle P_r(F_2^{(h)}), P_r(F_3^{(h)}) \rangle| dg \\ &\quad + |\langle P_r(F_2^{(h)}), P_r(F_3^{(h)}) \rangle| \\ [\epsilon\text{-mixing of right translation}] &\leq \epsilon \|F_2^{(h)}\|_{L^2} \|F_3^{(h)}\|_{L^2} + \left| \int_G F_2^{(h)}(x) dx \int_G F_3^{(h)}(x) dx \right| \\ \left[ \|F_2^{(h)}\|_{L^\infty} \leq 1 \text{ and } \|F_3^{(h)}\|_{L^\infty} = \|f_3\|_{L^\infty}^2 \leq 4 \right] &\leq 4\epsilon + \left| \int_G F_3^{(h)}(x) dx \right| = 4\epsilon + |\langle f_3, (h \cdot_c f_3) \rangle|. \end{aligned}$$

It remains to integrate over  $h$  and use  $\epsilon$ -mixing of conjugation as well as the assumptions of Case (ii):

$$\int \int_G |\langle e_g, e_{gh} \rangle| dg dh = 4\epsilon + \int_G |\langle f_3, (h \cdot_c f_3) \rangle| dh \leq 4\epsilon + \epsilon \|f_3\|_{L^2}^2 = 4\epsilon + \epsilon = 5\epsilon.$$

Recalling that  $D$ -quasirandom groups are  $D^{-1/2}$ -mixing and applying the above theorem, we get [Aus13, Theorem 1] with a slightly better bound: □

**Corollary 25.** *Let  $D \geq 1$  and  $(G, \mu)$  be a  $D$ -quasirandom compact Hausdorff group. Then for any  $f_1, f_2, f_3 \in L^2(G, \mu)$  with  $\|f_1\|_{L^\infty}, \|f_2\|_{L^\infty}, \|f_3\|_{L^\infty} \leq 1$ ,*

$$\int_G \left| \int_G f_1(x) (g \cdot_l f_2)(x) (g \cdot_c f_3)(x) dx - \int_G f_1(x) P_l(f_2)(x) P_c(f_3)(x) dx \right| dg \leq 4D^{-1/4}.$$

**Acknowledgements.** This paper owes a great deal to Sławek Solecki: his short note explaining the main result in [BT13], as well as generalizing Proposition 3 of [BT13] to arbitrary actions of quasirandom groups on finite sets, was what cleared things up for me and made me look from the right angle. Also, many thanks to Evgeny Gordon for useful conversations and comments; in particular, for bringing my attention to the fact that the notion of probability groups had already been defined in [Wei65], as well as for pointing out that the Loeb measure is *countably additive* even when defined on an ultraproduct of *finitely additive* probability spaces. Further thanks to Julien Melleray for pointing out an error in my original definitions of probability groups and their actions. Finally, thanks to Tim Austin for allowing me to use his argument in the quantitative van der Corput lemma, as well as for comments and corrections.

## REFERENCES

- [Aus13] T. Austin, *Quantitative equidistribution for certain quadruples in quasirandom groups*, preprint (2013).
- [CKTD12] C. T. Clinton, A. S. Kechris, and R. D. Tucker-Drob, *Ultraproducts of measure preserving actions and graph combinatorics*, *Ergodic Theory and Dynamical Systems* **33** (2012), no. 2, 334-374.
- [BM07] V. Bergelson and R. McCutcheon, *Central sets and a noncommutative Roth theorem*, *American Journal of Mathematics* **129** (2007), no. 5, 1251-1275.
- [BG04] V. Bergelson and A. Gorodnik, *Weakly mixing group actions: a brief survey and an example*, *Modern dynamical systems and applications*, Cambridge Univ. Press (2004), 3-25.
- [BT13] V. Bergelson and T. Tao, *Multiple recurrence in quasirandom groups*, preprint (2013).
- [Fur81] H. Furstenberg, *Recurrence in Ergodic Theory and Combinatorial Number Theory*, Princeton University Press, 1981.
- [Gow08] W. T. Gowers, *Quasirandom groups*, *Combin. Probab. Comput.* **17** (2008), no. 3, 363-387.
- [TD13] R. D. Tucker-Drob, *PhD Thesis: Descriptive set theory and the ergodic theory of countable groups*, California Institute of Technology, 2013.
- [Wei65] A. Weil, *L'intégration dans les Groupes Topologiques et Ses Applications*, 2nd ed., Publications de l'Institut de Mathématique de l'Université de Strasbourg, vol. 4, Hermann, 1965.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN, IL, 61801, USA  
*E-mail address:* anush@illinois.edu